mml-Analytic Geometry

Norms

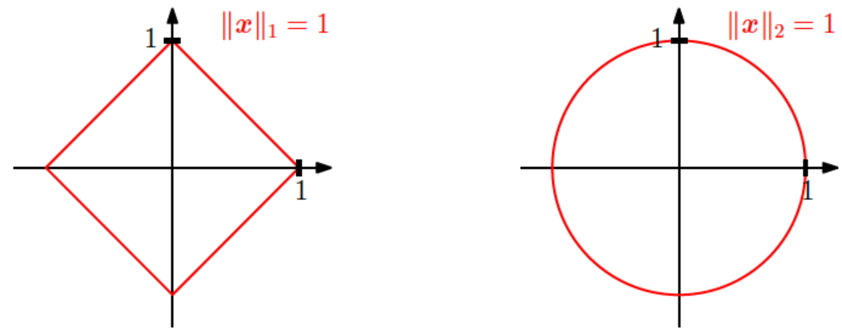
The *norms* are the length of the vectors.

> A norm on a vector space V is a function:
>
> $$\begin{aligned} \| \cdot \| &: V \to \mathbb{R}, \\ x &\mapsto \|x\| \end{aligned}$$

And for all $\lambda \in \mathbb{R}$ and $x, y \in V$, there are:

- *Absolutely homogeneous:* $\|\lambda x\| = |\lambda| \|x\|$
- *Positive definite*: $\|x\| \geq 0 \text{ and } \|x\| = 0 \iff x = 0$
- *Triangle inequality*: $\|x + y\| \leq \|x\| + \|y\|$

> **Manhattan Norm (l_1 *norm*)**
>
> The Manhattan norm is:
>
> $$\|x\|_1 := \sum_{i=0}^{n} |x_i|$$
>
> **Euclidean Norm (l_2 *norm*)**
>
> The Euclidean norm is:
>
> $$\|x\|_2 := \sqrt{\sum_{i=0}^{n} x_i^2} = \sqrt{x^T x}$$



Inner Products

Dot product

A particular type of inner product, the *scalar product/dot product*, which is:

$$x^T y = \sum_{i=0}^{n} x_i y_i$$

General Inner Products

Bilinear mapping

> A bilinear mapping Ω is a mapping with 2 arguments, and it's linear in each argument.
>
> each argument, i.e., when we look at a vector space V then it holds that for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V, \ \lambda, \psi \in \mathbb{R}$ that
>
> $$\Omega(\lambda \boldsymbol{x} + \psi \boldsymbol{y}, \boldsymbol{z}) = \lambda \Omega(\boldsymbol{x}, \boldsymbol{z}) + \psi \Omega(\boldsymbol{y}, \boldsymbol{z}) \tag{3.6}$$
> $$\Omega(\boldsymbol{x}, \lambda \boldsymbol{y} + \psi \boldsymbol{z}) = \lambda \Omega(\boldsymbol{x}, \boldsymbol{y}) + \psi \Omega(\boldsymbol{x}, \boldsymbol{z}) \, . \tag{3.7}$$
>
> Here Ω is linear in 1st argument in 3.6, and 2nd argument in 3.7.

> Symmetric and positive definite of bilinear mapping

Definition 3.2. Let V be a vector space and $\Omega : V \times V \to \mathbb{R}$ be a bilinear mapping that takes two vectors and maps them onto a real number. Then

- Ω is called *symmetric* if $\Omega(\boldsymbol{x}, \boldsymbol{y}) = \Omega(\boldsymbol{y}, \boldsymbol{x})$ for all $\boldsymbol{x}, \boldsymbol{y} \in V$, i.e., the order of the arguments does not matter.
- Ω is called *positive definite* if

$$\forall \boldsymbol{x} \in V \backslash \{\boldsymbol{0}\} : \Omega(\boldsymbol{x}, \boldsymbol{x}) > 0 \,, \quad \Omega(\boldsymbol{0}, \boldsymbol{0}) = 0 \,. \tag{3.8}$$

General linear products

Definition 3.3. Let V be a vector space and $\Omega : V \times V \to \mathbb{R}$ be a bilinear mapping that takes two vectors and maps them onto a real number. Then

- A positive definite, symmetric bilinear mapping $\Omega : V \times V \to \mathbb{R}$ is called an *inner product* on V. We typically write $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ instead of $\Omega(\boldsymbol{x}, \boldsymbol{y})$.
- The pair $(V, \langle \cdot, \cdot \rangle)$ is called an *inner product space* or (real) *vector space with inner product*. If we use the dot product defined in (3.5), we call $(V, \langle \cdot, \cdot \rangle)$ a *Euclidean vector space*.

Symmetric, Positive Definite Matrices

Consider an $n-$dimensional vector space V with an inner product $\langle \cdot, \cdot \rangle$ and an ordered basis $B = (b_1, \ldots, b_n)$ of V.

Since any vectors $x, y \in V$ can be written as linear combinations of the basis vectors so that $x = \sum_{i=1}^{n} \varphi_i b_i \in V$ and $y = \sum_{j=1}^{n} \lambda_j b_j \in V$.

Due to the bi-linearity of the inner product:

$$\langle x, y \rangle = \langle \sum_{i=1}^{n} \varphi_i b_i, \sum_{j=1}^{n} \lambda_j b_j \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} \varphi_i \langle b_i, b_j \rangle \lambda_j = \hat{x}^T A \hat{y}$$

where $A_{ij} := \langle b_i, b_j \rangle$and $\hat{x}, \hat{y}$ are the coordinates of x and y with respect to the basis B.

- Symmetry: A is symmetric
- Positive Definiteness: $\forall x \in V \setminus \{0\} : x^T A x > 0, \ (3.11)$

> Symmetric, Positive Definite Matrix
>
> **Definition 3.4** (Symmetric, Positive Definite Matrix). A symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ that satisfies (3.11) is called *symmetric, positive definite*, or just *positive definite*. If only $\geqslant$ holds in (3.11), then $\boldsymbol{A}$ is called *symmetric, positive semidefinite*.

If A is symmetric, positive definite, then

$$\langle x, y \rangle = \hat{x}^T A \hat{y}$$

defines an inner product with respect to an ordered basis B, where $\hat{x}$ and $\hat{y}$ are the coordinate representations of $x, y \in V$ with respect to B.

> Properties
>
> - The null space (kernel) of $\boldsymbol{A}$ consists only of $\mathbf{0}$ because $\boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} > 0$ for all $\boldsymbol{x} \neq \mathbf{0}$. This implies that $\boldsymbol{A}\boldsymbol{x} \neq \mathbf{0}$ if $\boldsymbol{x} \neq \mathbf{0}$.
> - The diagonal elements a_{ii} of $\boldsymbol{A}$ are positive because $a_{ii} = \boldsymbol{e}_i^\top \boldsymbol{A} \boldsymbol{e}_i > 0$, where $\boldsymbol{e}_i$ is the ith vector of the standard basis in $\mathbb{R}^n$.

Lengths and Distances

Inner products and norms are closely related. Mostly, any inner product can induces a norm.

$$\|x\| := \sqrt{\langle x, x \rangle}$$

But not every norm is induced by an inner product, like the Manhattan norm.

> *Remark*
>
> Cauchy-Schwarz Inequality: for an inner product vector space, the induced norm $\|\cdot\|$ satisfies

> $$|\langle \boldsymbol{x}, \boldsymbol{y}\rangle| \leq \|\boldsymbol{x}\|\|\boldsymbol{y}\|$$

Distance and Metric

Consider an inner product space $(V, \langle\cdot,\cdot\rangle)$. Then

$$d(x, y) := \|x - y\| = \sqrt{\langle x - y, x - y\rangle}$$

is called the *distance* between x and y for $x, y \in V$.

The *distance* is called *Euclidean distance* if use doc product as the inner product.

The mapping

$$\begin{aligned} d : V \times V &\to \mathbb{R} \\ (x, y) &\mapsto d(x, y) \end{aligned}$$

is called a *metric*.

> A metric d satisfies the following:
>
> 1. d is *positive definite*, i.e., $d(\boldsymbol{x}, \boldsymbol{y}) \geqslant 0$ for all $\boldsymbol{x}, \boldsymbol{y} \in V$ and $d(\boldsymbol{x}, \boldsymbol{y}) = 0 \iff \boldsymbol{x} = \boldsymbol{y}$.
> 2. d is *symmetric*, i.e., $d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{y}, \boldsymbol{x})$ for all $\boldsymbol{x}, \boldsymbol{y} \in V$.
> 3. *Triangle inequality:* $d(\boldsymbol{x}, \boldsymbol{z}) \leqslant d(\boldsymbol{x}, \boldsymbol{y}) + d(\boldsymbol{y}, \boldsymbol{z})$ for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$.

Angles and Orthogonality

Angles

Apart from the distance, inner products also capture the geometry of a vector space by defining the angle ω between two vectors. The angle shows how similar two vectors' orientations are.

Use the Cauchy-Schwarz inequality to define angles

$$-1 \leq \frac{\langle x, y \rangle}{\|x\|\|y\|} \leq 1$$

therefore, there exists a unique $w \in [0, \pi]$ has

$$cos\omega = \frac{\langle x, y \rangle}{\|x\|\|y\|}$$

Orthogonality

Two vectors x and y are *orthogonal* if and only if $\langle x, y \rangle = 0$ and written as $x \perp y$. If additionally $\|x\| = 1 = \|y\|$, the vectors are *orthonormal* .

> Vectors that are orthogonal with respect to one inner product do not have to be orthogonal with respect to a different inner product.
>
> $0-$vector is orthogonal to every vector in the vector space.

Orthogonal Matrix

A square matrix A is an *orthogonal matrix* if and only if its columns are orthonormal so that

$$AA^T = I = A^T A$$

which implies that

$$A^{-1} = A^T$$

the inverse is obtained by simply transposing the matrix.

Transformations by orthogonal matrices are special because the length of a vector x is not changed when transforming it using an orthogonal matrix A.

For dot product, we obtain

$$\|\boldsymbol{Ax}\|^2 = (\boldsymbol{Ax})^\top(\boldsymbol{Ax}) = \boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{Ax} = \boldsymbol{x}^\top\boldsymbol{Ix} = \boldsymbol{x}^\top\boldsymbol{x} = \|\boldsymbol{x}\|^2 \; .$$

The angle between any two vectors x, y is also unchanged when transforming both of them using an orthogonal matrix A.

For dot product, we also obtain

$$\cos\omega = \frac{(\boldsymbol{Ax})^\top(\boldsymbol{Ay})}{\|\boldsymbol{Ax}\|\,\|\boldsymbol{Ay}\|} = \frac{\boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{Ay}}{\sqrt{\boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{Axy}^\top\boldsymbol{A}^\top\boldsymbol{Ay}}} = \frac{\boldsymbol{x}^\top\boldsymbol{y}}{\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|}$$

Orthonormal Basis

An orthonormal basis is a special basis when the basis vectors are orthogonal to each other and their length are 1.

Consider an $n-$dimensional vector space V and a basis $\{b_1, \ldots, b_n\}$ of V. If

$$\langle b_i, b_j \rangle = 0 \text{ for i} \neq \text{j} \tag{1}$$
$$\langle b_i, b_j \rangle = 1 \tag{2}$$

for all $i, j = 1, \ldots, n$ then the basis is called an *orthonormal basis* (ONB).

🗣 If only (1), the basis is called an *orthogonal basis*.

Orthogonal Complement

Consider a D-dimensional vector space V and an M-dimensional subspace $U \subseteq V$. Then its *orthogonal complement* $U^\perp$ is a $(D - M)$-dimensional subspace of V and contains all vectors in V that are orthogonal to every vector in U.

And $U \cap U^\perp = \{0\}$ so that any vector $x \in V$ can be decomposed into

$$x = \sum_{m=1}^{M} \lambda_m b_m + \sum_{j=1}^{D-M} \varphi_j b_j^\perp$$

Therefore, the orthogonal complement can also be used to describe a plane in a three-dimensional vector space.

Generally, orthogonal complements can be used to describe hyperplanes in n-dimensional vector and affine spaces.

Inner Product of Functions

The inner product before are about the finite-dimensional vectors. Now consider a vector $x \in \mathbb{R}$ as a function with n function values.

- vectors with an infinite number of entries → countable infinite
- continuous-valued functions → uncountable infinite

An inner product of 2 functions $u : \mathbb{R} \to \mathbb{R}$ and $v : \mathbb{R} \to \mathbb{R}$ can be defined as the definite integral

$$\langle u, v \rangle := \int_a^b u(x)v(x)dx.$$

- if the result is 0, the functions u and v are orthogonal
- inner product on functions may diverge

Orthogonal Projections

Projections are important in ML and some other areas. Compressing data in high-dimension to lower dimension space and to minimise the compression loss, we need to find the most informative dimensions in the data.

> Projection
>
> **Definition 3.10** (Projection). Let V be a vector space and $U \subseteq V$ a subspace of V. A linear mapping $\pi : V \to U$ is called a *projection* if $\pi^2 = \pi \circ \pi = \pi$.

Since linear mappings can be expressed by transformation matrices, the preceding definition applies to a special kind of transformation matrices, the *projection matrices* P_π, which exhibit the property that $P_\pi^2 = P_\pi$.

In the following parts, assume the dot product $\langle x, y\rangle = x^T y$ as the inner product.

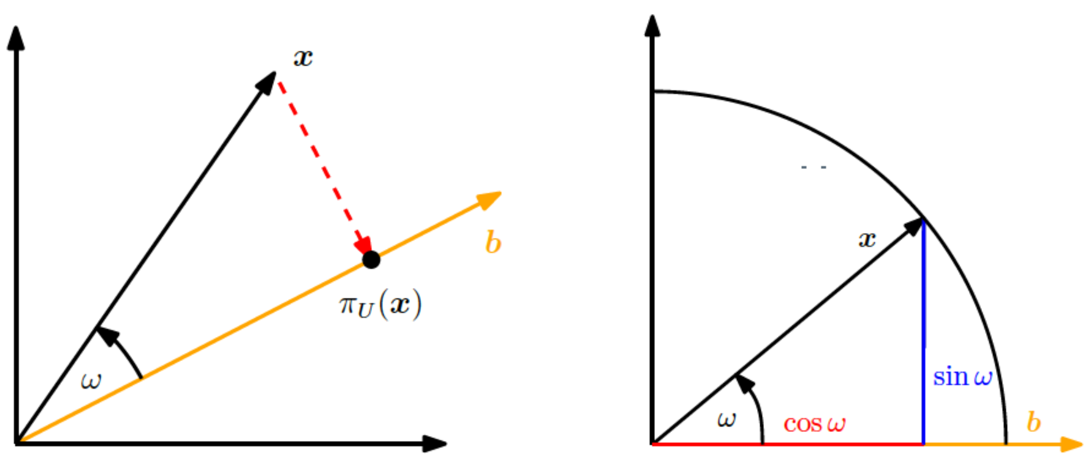
Projection onto 1-d subspaces

Assume a line through the origin with a basis vector $b \in \mathbb{R}^n$. This line is a one-dimensional subspace $U \in \mathbb{R}^n$ spanned by b.

When project $x \in \mathbb{R}^n$ onto U, the vector(projection) $\pi_U(x)$ is closest to x.

Properties of the projection

- The projection $\pi_U(x)$ is closest to x, which means the distance $\|x - \pi_u(x)\|$ is minimal.
    - $\pi_U(x) - x$ is orthogonal to U
    - $\langle \pi_U(x) - x, b\rangle = 0$
- The projection $\pi_U(x)$ of x onto U must be an element of U and a multiple of the basis vector b that spans U.
    - hence, $\pi_U(x) = \lambda b$.



(a) Projection of $x \in \mathbb{R}^2$ onto a subspace U with basis vector b.

(b) Projection of a two-dimensional vector x with $\|x\| = 1$ onto a one-dimensional subspace spanned by b.

Find coordinate λ

The orthogonality condition,

$$\langle x - \pi_U(x), b\rangle = 0 \overset{\pi_U(x)=\lambda b}{\iff} \langle x - \lambda b, b\rangle = 0$$

since the bi-linearity of inner product

$$\langle x, b\rangle - \lambda\langle b, b\rangle = 0 \iff \lambda = \frac{\langle x, b\rangle}{b, b} = \frac{\langle b, x\rangle}{||b||^2}$$

choose $\langle \cdot, \cdot \rangle$ to be the dot product

$$\lambda = \frac{b^T x}{b^T b} = \frac{b^T x}{||b||^2}$$

If $||b|| = 1$, the coordinate λ of the projection is $b^T x$.

> For general inner product, $\lambda = \langle x, b\rangle$ if $||b|| = 1$.

Find the projection point $\pi_U(x) \in U$

since $\pi_U(x) = \lambda b$, with the result above

$$\pi_U(x) = \lambda b = \frac{\langle x, b\rangle}{||b||^2} b = \frac{b^T x}{||b||^2} b$$

Also the length of $\pi_U(x)$ is

$$||\pi_U(x)|| = ||\lambda b|| = |\lambda| ||b||$$

If use the dot product as an inner product

$$||\pi_U(x)|| = \frac{b^T x}{||b||^2} ||b|| = |cos\omega| ||x|| ||b|| \frac{||b||}{||b||^2} = |cos\omega| ||x||$$

w is the angle between x and b.

If $||x|| = 1$, then x lies on the unit circle, shown as (b) on the figure above.

Find the projection matrix P_π

Since projection is a linear mapping, there's a projection matrix P_π and $\pi_U(x) = P_\pi x$, with the dot product

$$\pi_U(x) = \lambda b = b\lambda = b\frac{b^T x}{||b||^2} = \frac{bb^T}{||b||^2}x$$

therefore,

$$P_\pi = \frac{bb^T}{||b||^2}$$

bb^T is a symmetric matrix and $||b||^2 = \langle b, b\rangle$ is a scalar.

Projection onto general subspaces

Assume $(b_1, \ldots, b_n)$ is an ordered basis of U, any projection $\pi_U(x)$ onto U is an element of U. Therefore, they can be represented as linear combinations of the basis vectors of U, $\pi_U(x) = \sum_{i=1}^m \lambda_i b_i$.

Find coordinates $\lambda_1, \ldots, \lambda_m$

$$\pi_U(x) = \sum_{i=1}^m \lambda_i b_i = B\lambda,$$
$$B = [b_1, \ldots, b_m], \quad \lambda = [\lambda_1, \ldots, \lambda_m]$$

since orthogonal

$$\langle b_i, x - \pi_U(x)\rangle = b_i^T(x - \pi_U(x)) = 0$$

also can be written as

$$b_i^T(x - B\lambda) = 0$$

therefore,

$$\begin{bmatrix} \boldsymbol{b}_1^\top \\ \vdots \\ \boldsymbol{b}_m^\top \end{bmatrix} \begin{bmatrix} \boldsymbol{x} - \boldsymbol{B\lambda} \end{bmatrix} = \boldsymbol{0} \iff \boldsymbol{B}^\top(\boldsymbol{x} - \boldsymbol{B\lambda}) = \boldsymbol{0}$$
$$\iff \boldsymbol{B}^\top\boldsymbol{B\lambda} = \boldsymbol{B}^\top\boldsymbol{x}\,.$$

since the basis is linearly independent, B^TB is regular and can be inverted,

$$\lambda = (B^TB)^{-1}B^Tx$$

$(B^TB)^{-1}B^T$ is also the *pseudo-inverse* of B.

Find the projection $\pi_U(x)$

$$\pi_U(x) = B\lambda = B(B^TB)^{-1}B^Tx$$

Find the projection matrix P_π

$$P_\pi x = \pi_U(x),$$
$$P_\pi = B(B^TB)^{-1}B^T$$

- The projection $\pi_U(x)$ is still vector in $\mathbb{R}^n$ although they lie in an m-dimensional subspace

Projections can help find *approximate solution* to the linear equation that can't be solved exactly by finding the vector in the subspace spanned by columns of A that's closest to b,

- compute the orthogonal projection of b onto the subspace spanned by the columns of A.

the solution is called *least-squares solution*.

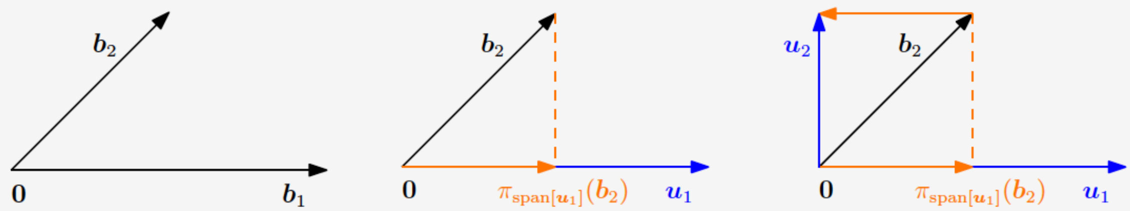
Gram-Schmidt Orthogonalisation

Gram-Schmidt is a method that transform any basis $(b_1, \ldots, b_m)$ of an n-dimensional vector space V into an orthogonal/orthonormal basis $(u_1, \ldots, u_m)$.

> Definition
>
> $\mathrm{span}[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n]$. The *Gram-Schmidt orthogonalization* method iteratively constructs an orthogonal basis $(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)$ from any basis $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ of V as follows:
>
> $$\boldsymbol{u}_1 := \boldsymbol{b}_1 \tag{3.67}$$
> $$\boldsymbol{u}_k := \boldsymbol{b}_k - \pi_{\mathrm{span}[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{k-1}]}(\boldsymbol{b}_k)\,, \quad k = 2, \ldots, n\,. \tag{3.68}$$

> Example

Example 3.12 (Gram-Schmidt Orthogonalization)



(a) Original non-orthogonal basis vectors $\boldsymbol{b}_1, \boldsymbol{b}_2$.

(b) First new basis vector $\boldsymbol{u}_1 = \boldsymbol{b}_1$ and projection of $\boldsymbol{b}_2$ onto the subspace spanned by $\boldsymbol{u}_1$.

(c) Orthogonal basis vectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2 = \boldsymbol{b}_2 - \pi_{\mathrm{span}[\boldsymbol{u}_1]}(\boldsymbol{b}_2)$.

Consider a basis $(\boldsymbol{b}_1, \boldsymbol{b}_2)$ of $\mathbb{R}^2$, where

$$\boldsymbol{b}_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \quad \boldsymbol{b}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}; \tag{3.69}$$

see also Figure 3.2(a). Using the Gram-Schmidt method, we construct an orthogonal basis $(\boldsymbol{u}_1, \boldsymbol{u}_2)$ of $\mathbb{R}^2$ as follows (assuming the dot product as the inner product):

$$\boldsymbol{u}_1 := \boldsymbol{b}_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \tag{3.70}$$

$$\boldsymbol{u}_2 := \boldsymbol{b}_2 - \pi_{\mathrm{span}[\boldsymbol{u}_1]}(\boldsymbol{b}_2) \overset{(3.45)}{=} \boldsymbol{b}_2 - \frac{\boldsymbol{u}_1 \boldsymbol{u}_1^\top}{\|\boldsymbol{u}_1\|^2} \boldsymbol{b}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{3.71}$$